

Lebih Lanjut Dengan Regresi Logistik

Oleh: Syahrudji Naseh
Puslit Ekologi Kesehatan, Badan Litbangkes

DALAM tulisan yang berjudul "Beberapa Model Statistik dalam Penelitian Kesehatan / Kedokteran" (Media Litbangkes Vol.II No. 2, 1992) diperkenalkan salah satu metoda analisis data yaitu regresi logistik ("logistic regression"). Pada tulisan tersebut, diberikan juga contoh penggunaannya secara selintas. Untuk mengenal lebih dalam tentang regresi logistik, tulisan ini mencoba mengulas interpretasi dan cara perhitungannya. Akan tetapi, karena persamaan regresi logistik sangat erat hubungannya dengan perhitungan "Relative Risk" (RR) dan "Odds Ratio" (OR), maka hal ini akan dibahas terlebih dahulu.

"RELATIVE RISK" (RR) DAN "ODDS RATIO" (OR)

RR adalah salah satu data kuantitatif yang mengukur hubungan antara variabel bebas dan variabel tidak bebas di dalam analisis tabel 2 x 2 (lihat Tabel 1). Status variabel bebas merupakan variabel paparan ("exposure status") yang dibagi dalam 2 kategori yaitu terpapar ("ya") atau tidak terpapar ("tidak"). Sedangkan status variabel tidak bebas merupakan variabel penyakit ("disease status") yang juga terbagi 2 kategori yaitu "ya dan "tidak". Sebagai catatan, status penyakit disini bukan saja berarti mengidap atau tidak mengidapnya seorang pasien atau responden atau unit analisis lain terhadap suatu penyakit. Akan tetapi dapat juga merupakan mati dan

hidupnya bayi (Tabel 2) ataupun baik dan buruknya gizi balita (Tabel 3).

Tabel 1. Tabel 2 x 2 serta Rumus Perhitungan RR dan OR

ADA PAPARAN	ADA PENYAKIT		
	Ya	Tidak	Jumlah
Ya	A	B	M1
Tidak	C	D	M2

$$RR = (A/M1)/(C/M2)$$

$$OR = AD/BC$$

Tabel 2. Tabel 2 x 2 untuk menghitung RR dan OR pada kematian bayi

UMUR IBU	STATUS BAYI		
	Mati	Hidup	Jumlah
15-19 tahun	29	48	77
20-34 tahun	134	311	445

Sumber: SKRT 1986

$$RR = (29/77)/(134/445) = 0,38/0,30 = 1,27$$

$$OR = (29 \times 311)/(48 \times 134) = 1,40$$

Tabel 3. Tabel 2 x 2 untuk menghitung RR dan OR pada gizi balita

DESA/KOTA	GIZI BALITA		
	Buruk	Baik	Jumlah
Perkotaan	421	579	1000
Pedesaan	616	384	1000

Sumber: Fiktif

$$RR = (421/1000)/(616/1000) = 0,68$$

$$OR = (421 \times 384)/(579 \times 616) = 0,45$$

Dari Tabel 1 dapat dilihat bahwa A adalah banyaknya unit analisis yang terpapar dan terkena penyakit. B adalah banyaknya unit analisis yang terpapar tetapi tidak terkena penyakit. C adalah banyaknya unit analisis yang tidak terpapar tetapi terkena penyakit. D adalah banyaknya unit analisis yang tidak terpapar dan tidak terkena penyakit. Sedangkan M1 dan M2 masing-masing merupakan jumlah unit analisis yang terpapar (A+B) dan yang tidak terpapar (C+D). Rumus RR dan OR dapat dilihat di bawah Tabel 1 ini. RR mengukur berapa kali besarnya suatu penyakit akan menimpa unit analisis yang terpapar dibandingkan dengan kelompok yang tidak terpapar. Jika RR tidak sama dengan 1, maka ada hubungan antara "study factor" dan "risk of disease". Hubungan itu disebut hubungan positif bila RR lebih besar dari 1. Sebaliknya, bila RR kurang dari 1 disebut hubungan negatif.

Untuk penelitian yang bersifat penelitian prospektif/kohor dan "cross-sectional" yang sampelnya besar, digunakan ukuran RR. Akan tetapi, untuk penelitian kasus-kelola ("case-control") yang biasanya mengambil sampel relatif kecil, digunakan ukuran OR yang merupakan estimasi (nilai perkiraan) untuk RR. Tentang jenis-jenis penelitian ini, lebih lanjut dapat dibaca umpamanya tulisan yang berjudul "Keunggulan dan Keterbatasan Beberapa Metode Penelitian Kesehatan / Kedokteran" (Media Litbangkes Vol. III No. 1, 1993).

Tabel 2 memperlihatkan bahwa dari 77 bayi yang dilahirkan oleh ibu yang berumur 15 sampai 19 tahun, 29 telah meninggal pada waktu wawancara SKRT (Survei Kesehatan Rumah Tangga) 1986. Sementara itu, dari 445 bayi yang dilahirkan oleh ibu yang berumur antara 20 sampai dengan 34 tahun, 134 telah meninggal dan sisanya masih hidup pada saat wawancara. Dengan bantuan

kalkulator, dapat dihitung besarnya faktor risiko (RR dan OR) yang besarnya masing-masing 1,27 dan 1,40. Karena yang digunakan untuk analisis lanjut SKRT 1986 ini pendekatan kasus-kelola, lebih baik dipakai OR yang besarnya 1,40. Ini berarti, bahwa risiko kematian bayi 1,40 kali lebih besar pada bayi yang dilahirkan ibu yang berumur muda (15 - 19 tahun) dibandingkan dengan bayi yang dilahirkan oleh ibu dengan umur ideal (20 - 34 tahun).

Pada Tabel 3, 2000 sampel balita diamati dengan metode kohor / prospektif. Sampel itu masing-masing berjumlah 1000 balita dari desa dan kota. Variabel tidak bebas merupakan status gizi balita yaitu baik dan buruk. Sedangkan variabel tidak bebas adalah tempat tinggal balita tersebut yang dikategorikan sebagai desa dan kota. Ternyata, 421 balita yang tinggal di kota bergizi buruk dan 579 bergizi baik. Sedangkan 384 balita yang tinggal di desa bergizi baik dan 384 bergizi buruk. Maka $RR = 0,68$ dan $OR = 0,45$. Dengan demikian, dapat disimpulkan bahwa risiko balita yang tinggal di kota untuk mendapat gizi buruk 0,68 kali dibandingkan bila ia tinggal di desa (yang dipakai disini RR).

Dari analisis tabel 2x2 dapat dihitung hubungan antara satu variabel tidak bebas dan satu variabel bebas. Padahal pada kenyataannya banyak sekali variabel bebas yang mempengaruhi variabel tidak bebas ini. Pada analisis lanjut SKRT 1986, ada 4 faktor yang diduga mempunyai hubungan dengan variabel tidak bebas kematian bayi, yaitu faktor-faktor biomedis ibu, pelayanan kesehatan, sosial ekonomi dan lingkungan. Faktor biomedis ibu terdiri dari umur ibu, paritas ("parity"), jarak kelahiran ("birth interval"), umur kehamilan pada waktu melahirkan, komplikasi sewaktu hamil dan komplikasi sewaktu bersalin. Tabel 4 merupakan ringkasan hasil

perhitungan OR dari tabel 2 x 2 disamping ditampilkan pula proporsinya. Kategori yang OR nya sama dengan 1 merupakan kategori ideal atau "reference category", yang biasanya diletakkan pada baris ke dua tabel 2 x 2. Ringkasan perhitungan OR dari Tabel 2 umpamanya, dapat dilihat pada baris 1 sampai dengan 3 Tabel 4. Dalam hal ini, kategori ideal umur ibu adalah 20 sampai 34 tahun dengan OR sama dengan 1.

Tabel 4. Distribusi faktor Biomedis ibu pada kematian bayi

	BAYI				OR
	MATI		HIDUP		
	#	%	#	%	
UMUR IBU					
15-19 tahun	29	16,0	48	12,4	1,40
20-34 tahun	134	74,0	311	80,2	1
>34 tahun		9,9	29	7,5	1,44
	18				
PARITAS					
1 kali	53	29,3	67	17,3	2,51
2-3 kali	64	35,4	203	52,3	1
4-6 kali	42	23,2	92	23,7	1,45
>6 kali	22	12,2	26	6,7	2,68
JARAK KELAHIRAN					
9-17 bulan	17	13,4	126	39,8	0,19
18-23 bulan	18	14,1	71	22,4	0,36
24-35 bulan	42	33,1	59	18,6	1
>35 bulan	50	39,4	61	19,2	1,15
UMUR KEHAMILAN					
28-36 minggu	27	15,0	31	8,0	2,03
>36 minggu	153	86,0	356	92,0	1
KOMPLIKASI KEHAMILAN					
Tidak ada	132	72,9	306	78,9	1
Ada	49	27,1	82	21,1	1,39
KOMPLIKASI PERSALINAN					
Tidak ada	147	81,2	318	82,0	1
Ada	34	18,8	70	18,0	1,05

Sumber: SKRT 1986

REGRESI LOGISTIK

Analisis tabel 2x2 dapat melihat adanya hubungan antara satu variabel tidak bebas dengan beberapa variabel bebas satu per satu. Ini disebut juga sebagai analisis "bivariate". Dengan bantuan regresi logistik, dapat dilakukan analisis "multivariate". Analisis ini melihat hubungan antara satu variabel tidak bebas dan beberapa variabel bebas bersama-sama.

Persamaan regresi logistik dapat ditulis sebagai:

$$f(y) = 1/[1 + \exp\{-(a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n)\}]$$

(persamaan 1)

atau dapat disederhanakan sebagai:

$$\text{logit } f(y) = \ln [f(y)/(1 - f(y))] \quad (\text{persamaan 2})$$

$$= a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

(persamaan 3)

atau

$$f(y)/(1 - f(y)) = \exp(a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n)$$

(persamaan 4).

Dalam hal ini, $f(y)$ adalah probabilitas adanya suatu penyakit pada responden, \exp adalah bilangan alam atau bilangan e yang besarnya 2,718..., $x_1, x_2, x_3, \dots, x_n$ adalah variabel bebas ke 1, 2, 3, ..., n .

Bila dilakukan perhitungan dengan regresi logistik pada Tabel 4 untuk variabel bebas komplikasi kehamilan, maka persamaannya akan tertulis sebagai $\text{logit } f(y) = -0,6778 + 0,1629 \text{ KH}(1)$. Dalam hal ini, KH adalah komplikasi kehamilan. KH(1) adalah komplikasi kehamilan bila ada dan KH(-1) adalah komplikasi kehamilan bila tidak ada.

Untuk mencari OR, gunakan persamaan 4 untuk membandingkan yang ada komplikasinya (KH=1) dengan yang tidak ada komplikasinya (KH=-1), yaitu $[\exp\{-0,6778 + 0,1629(1)\}]/[\exp\{-0,6778 + 0,1629(-1)\}] = \exp(0,3258) = 1,3851 = 1,39$. Jadi sama

dengan hasil yang tertera pada Tabel 4 yang dihitung menggunakan analisis tabel 2x2, yaitu OR = 1,39. Ini berarti bahwa bayi yang lahir dari ibu yang mengalami komplikasi hamil mempunyai risiko untuk meninggal 1,39 kali lebih besar dari bayi yang lahir dari ibu yang tidak punya komplikasi hamil.

Bila selain variabel KH (komplikasi kehamilan) dimasukkan juga variabel KB (komplikasi bersalin), maka persamaan regresi logistiknya menjadi logit $f(y) = -0,6838 + 0,1649 KH(1) - 0,0111 KB(1)$. Dalam hal ini, KH(1) merupakan adanya komplikasi kehamilan, KH(-1) adalah tidak ada komplikasi kehamilan, KB(1) merupakan adanya komplikasi bersalin dan KB(-1) adalah tidak ada komplikasi bersalin. Dengan demikian, ada 4 kombinasi persamaan regresi logistik yaitu:

$$f(y)/\{1-f(y)\} = \exp\{-0,6838 + 0,1649(1) - 0,0111(1)\} \quad (\text{persamaan 5})$$

$$f(y)/\{1-f(y)\} = \exp\{-0,6838 + 0,1649(-1) - 0,0111(1)\} \quad (\text{persamaan 6})$$

$$f(y)/\{1-f(y)\} = \exp\{-0,6838 + 0,1649(1) - 0,0111(-1)\} \quad (\text{persamaan 7})$$

$$f(y)/\{1-f(y)\} = \exp\{-0,6838 + 0,1649(-1) - 0,0111(-1)\} \quad (\text{persamaan 8}).$$

Dalam hal ini, persamaan 5 berlaku untuk bayi yang lahir dari ibu yang mempunyai komplikasi hamil maupun bersalin. Persamaan 6 berlaku untuk bayi yang lahir dari ibu yang tidak mempunyai komplikasi hamil tetapi punya komplikasi bersalin. Persamaan 7 berlaku untuk bayi yang lahir dari ibu yang mempunyai komplikasi hamil tetapi tidak punya komplikasi bersalin. Persamaan 8 merupakan kategori ideal yaitu untuk bayi yang lahir dari ibu yang tidak punya komplikasi hamil maupun komplikasi bersalin. Jadi, untuk menghitung OR sebaiknya persamaan 5 sampai dengan persamaan dibandingkan dengan persamaan 8, sebagai berikut:

- ♦ Persamaan 5 : persamaan 8 = 1,36 : 1
- ♦ Persamaan 6 : persamaan 8 = 0,98 : 1
- ♦ Persamaan 7 : persamaan 8 = 1,39 : 1.

Ini berarti bahwa bayi yang lahir dari ibu yang mengalami komplikasi hamil dan bersalin mempunyai risiko untuk meninggal 1,36 kali lebih besar dari bayi yang lahir dari ibu yang tidak punya komplikasi hamil maupun bersalin. Demikian juga, bayi yang lahir dari ibu yang mengalami komplikasi bersalin saja mempunyai risiko untuk meninggal 0,98 kali lebih besar dari bayi yang lahir dari ibu yang tidak punya komplikasi hamil maupun bersalin.

Terakhir, bayi yang lahir dari ibu yang mengalami komplikasi hamil saja mempunyai risiko untuk meninggal 1,39 kali lebih besar dari bayi yang lahir dari ibu yang tidak punya komplikasi hamil maupun bersalin.

Untuk persamaan regresi logistik dengan lebih dari 2 variabel bebas, kombinasi yang terjadi akan lebih banyak.

Demikian juga, interpretasinya akan lebih sulit. Apalagi kalau beberapa variabel bebasnya bukan variabel diskrit / kategorikal atau merupakan interaksi dari 2 atau lebih variabel bebas.

DAFTAR PUSTAKA

1. Marija J. Norusis, SPSS/PC+ Advanced Statistics 4.0 for the IBM PC/XT/AT and PS/2, SPSS Inc., Chicago, 1990.
2. Schlesselman, J.J. Case-control Studies: Design, Conduct, Analysis. Oxford University Press, 1982.
3. W.J.Dixon, BMDP Statistical Software, University of California Press, California, 1985.
4. Zainul Bakri, et. al., Faktor-faktor yang Berperan dalam Kematian Bayi, Analisis Lanjut SKRT 1986, Makalah presentasi Badan Litbang Kesehatan tanggal 18 Nopember 1991. □